Ph.D. Synopsis

An Integrated Framework For Predicting Students' Academic Performance In Relation To The Affective And Cognitive State In The Learning Environment Using Multimodal Data

By

Amala Jayanthi.M

(Reg No. 19PHCSP001)

Supervisor

Dr.I.Elizabeth Shanthi

Professor, Department of Computer Science

Synopsis Submitted to

Avinashilingam Institute for Home Science and Higher Education for Women,

Coimbatore-43

In partial Fulfilment of the Requirement for the Degree of

Doctor of Philosophy in Computer Science

APRIL 2023

Introduction:

State of the Art:

Education is the process of developing knowledge, skills, attitudes, beliefs, and habits through various learning experiences. It is a way for people to learn to think critically, solve issues, communicate clearly, and gain a wide-ranging grasp of their surroundings. The main purpose of education is to provide people with the skills they need to live happy, productive lives and positively contribute to society. Teachers and students are the heart and soul of the educational system. Students develop their knowledge, skills, and behaviours through education and proper surveillance. Teachers play a critical role in shaping the future of society by equipping students with the knowledge, skills, and values they need to succeed in their personal and professional lives. Bloom's taxonomy has significantly impacted the educational system as it emphasizes the importance of developing higher-order thinking skills in students. According to Bloom's theory on educational activity, a person's capacity to learn is influenced by their mental (cognitive) and emotional (affective) domains. One of the emerging uses in education is educational data mining. It aids in a deeper comprehension of how students learn and interact with the process intended to raise the standard and financial viability of the educational system. This research focuses on the influence of students' Affective domain on their cognitive domain. This research proposes a recommender system for predicting student academic performance by associating the attributes of the affective domain, cognitive domain and academic performance. The 4 phases of the proposed research are Phase 1 – Quest-SA, A Evaluating Method for Closed-Ended Questionnaires using Sentiment Analysis by assigning polarity scores. Phase 2: Re-clust, Re-clustering and Cluster Validation based on mixed data clustering algorithm. Phase 3: ARMpred, An Efficient Prediction Based on Rule Generation for Clustered Data through Associative Classifier. Phase 4: Recommender Systems for Predicting Student academic performance. Quest-SA, Re-clust, and ARMpred's proposed algorithms' experimental results show superior performance than the existing models. It demonstrates a significant link between the mental and emotional measures that students prefer. The teaching profession can now better grasp their students' emotional conditions, give them the necessary instruction to advance their academic proficiency and also helps to succeed in their personal and professional lives.

Motivation

Education is a potent change agent that creates prosperous people, wholesome neighbourhoods, strong societies, vibrant economies, and a healthy planet. Every society's fundamental building component is education. Every person has a right to high-quality education and opportunities for lifelong learning. Education is a planned activity with certain goals, such as knowledge transmission or developing skills and character. Numerous ideas characterize education; Bloom's Taxonomy is one of them and is also the most well-known. Dr Benjamin Bloom, an educational psychologist, developed a system in 1956 to categorize learning objectives into many domains that motivate teachers to approach education holistically. Bloom classified learning activities into three domains: the cognitive domain (knowledge or mental skills), the affective domain (attitude or emotions), and the psychomotor or physical domain. Bloom emphasizes that a student grows knowledge worthy, emotionally and skillfully during learning. All three domains influence one another. The objective of education is attained when there is the highest development of the individual as a member with knowledge and positive behaviour traits that influence their self-growth and societal growth. This research studies the relationship between the cognitive domain and affective domain to help the tutor to bridge the gap between the chosen cognitive parameter and affective parameters using educational data mining.

The vast amount of data emerging in the educational sphere is the focus of the research field known as educational data mining. In the following forms of the discovering data as a data set, different processing methodologies are employed to derive an intriguing pattern from it, which can then be used for further analysis and decision-making for the re-modelling of learning patterns, improvement of teaching methods, as well as for the increase in quality of the educational system.

Emotional intelligence is considered the main affective attribute. Self Esteem, Self Determination, Self Efficacy, Well being, Positive Emotions and Negative Emotions, and Personality are second-level affective attributes to assess the affective domain. This research studies the association between the main attribute, second-level attributes and academic performance, the attribute for assessing the cognitive domain.

A criterion-reference model is a method for determining how well a student has met the course's objective. The evaluation is conducted using predetermined criteria. The findings are relationships between the student's performance and the specified standards. The grade is given

according to the criteria the student met concerning the requirements. CGPA is used to assess the student's academic performance. The academic performance is categorized as Excellent (CGPA: 9.00 - 9.99), Very Good (8.00 - 8.99), Good (7.00 - 7.99), Above Average (6.00 - 6.99), Average (5.00 - 5.99), Fail(<5.0).

Problem Definition

The wide concept of education includes gaining general knowledge, self-awareness, and skill development. Education is a crucial element for behaviour change, even while it is insufficient. As in Bloom's theory, the outcome of education is fulfilled only when the student is groomed in knowledge, Behavior/Emotions and skills. An effective teacher is the most important factor in a student's learning. This research proposes a recommender system for predicting student academic performance with the following findings,

- Impact of Emotional Intelligence on Self Esteem, Self Determination, Emotional Well-being, Self Efficacy, Personality Individually.
- Influence of Self Esteem, Self Determination, Emotional Well-being, Self Efficacy, and Personality on academic performance individually.

This system helps the teacher understand the students' affective state, which is lagging. By analyzing the gap between the chosen affective attributes, teachers can be emotional mentors so that they groom their students affectively strong to excel in academics.

Objectives of the Research

The research objective is to develop a model that predicts student academic performance by associating the attributes of the affective domain, cognitive domain and academic performance. The key objectives of this research as,

- To consider emotional intelligence as a main affective attribute.
- Select Self Esteem, Self Determination, Self Efficacy, Well being, Positive Emotions and Negative Emotions, and Personality as second-level affective attributes.
- To use closed-ended questionnaires for predicting student affective state.

- To create a sentiment analysis model for identifying student affective states by assigning a polarity value for each questionnaire scale.
- To create a mixed numerical and categorical data clustering algorithm.
- To generate proper rules from clustered data by developing a hybrid association rule mining algorithm.
- To develop a recommender system for predicting student academic performance using association rules mined.

Review of Literature

Educational data mining uses conventional data mining techniques for issues about education. It refers to applying data mining techniques to educational data, such as student data, academic records, exam results, participation in class, and the frequency of questions raised by students [1]. It has developed into a useful tool in recent years for predicting academic success, finding underlying knowledge in educational data, and enhancing the learning and teaching environment.

Cruz-Jesus et al. [2] used demographic factors, including age, gender, class participation, web access, computer ownership, and the number of courses taken, to predict student academic achievement. Many machine learning techniques were effective at forecasting student performance. A machine-learning model based on learning techniques, social support perception, motivation, socio-demographics, health status, and academic performance factors was proposed by Musso et al. [3]. The author used this model to forecast academic success and dropout rates. It has been determined that learning methodologies had the biggest impact on predicting GPA, whereas background knowledge had the most impact on predicting dropouts. To investigate the relationship between the data produced by the educational process in a mixed educational environment and academic performances, Hung et al. [4] employ machine learning and symmetry-based learning algorithms. The learning records students produced in a mixed learning environment were gathered for this study, including synchronous and asynchronous involvement, assignment submissions, and conversation in the online forum.

De Medio et al. [5] suggest a hybrid recommender strategy for an educational management system. Following a straightforward keyword-based query, it can search across many supported

standard-compliant learning object repositories and offer a prioritized list of learning objects. Furthermore, socially generated features demonstrate to the educator how the mentioned Learning Objects have been used in other courses. Baidada et al. [6] suggest combining content-based and collaborative filtering approaches to enhance the significance of recommendations in an online educational context.

A machine learning-based method for sentiment assessment of students' opinions on higher education institutions is presented by Toçolu et al. [7]. Chen et al. [8] present a novel event-based approach to handle the sentiment analysis process's difficulty in education. Rajput et al. [9] presented sentiment analysis using data collected from students after an instructor assessment session. KNIME's text processing capabilities were used to construct a pipeline for studying student emotions. This approach suggests utilizing an emotion score to label responses as positive, negative, or neutral. Nasim et al. [10] suggested combining lexicon-based techniques and machine learning to assess student input emotions. Textual feedback, typically provided after a course, offers helpful perceptions of the general quality of instruction and suggests feasible approaches to improve instructing techniques.

A clustering method for mixed data with a noise-filtered distribution centroid and an incremental weight alteration approach is provided by Li et al. [11]. D'Urso and Massari [12] suggest a fuzzy clustering model for data with diverse attributes. A variety of traits or qualities can be considered using the clustering methodology. The distance measure for numerous attributes is provided by aggregating the heterogeneity measures for each feature and using a weighting technique. Throughout the optimization phase, the weights are determined objectively. The weights in the clustering results signify the significance of each sort of attribute. Rodriguez et al. [13] propose a multi-partition clustering method incorporating Bayesian approaches to handle mixed data effectively.

Rajab [14] suggests Active Pruning Rules as a new approach for improving classifier performance, particularly prediction accuracy and decreasing rule redundancy. It preserves the rules in the classifier with actual data coverage and rejects any rule whose data occurrence has dropped below the permissible least occurrence. Pal et al. [15] present heuristic approaches for

efficient rule development and selection. It first uses the database to reduce the search space and then explicitly investigates the most powerful class association rules from the optimized database. Finally, Mattiev et al. [16] propose a new associative classifier that employs agglomerative hierarchical clustering as a post-processing step to minimize the number of rules.

Methodology

The overall workflow representation of the framework proposed is illustrated in Figure 1. This research work has been implemented in four phases.

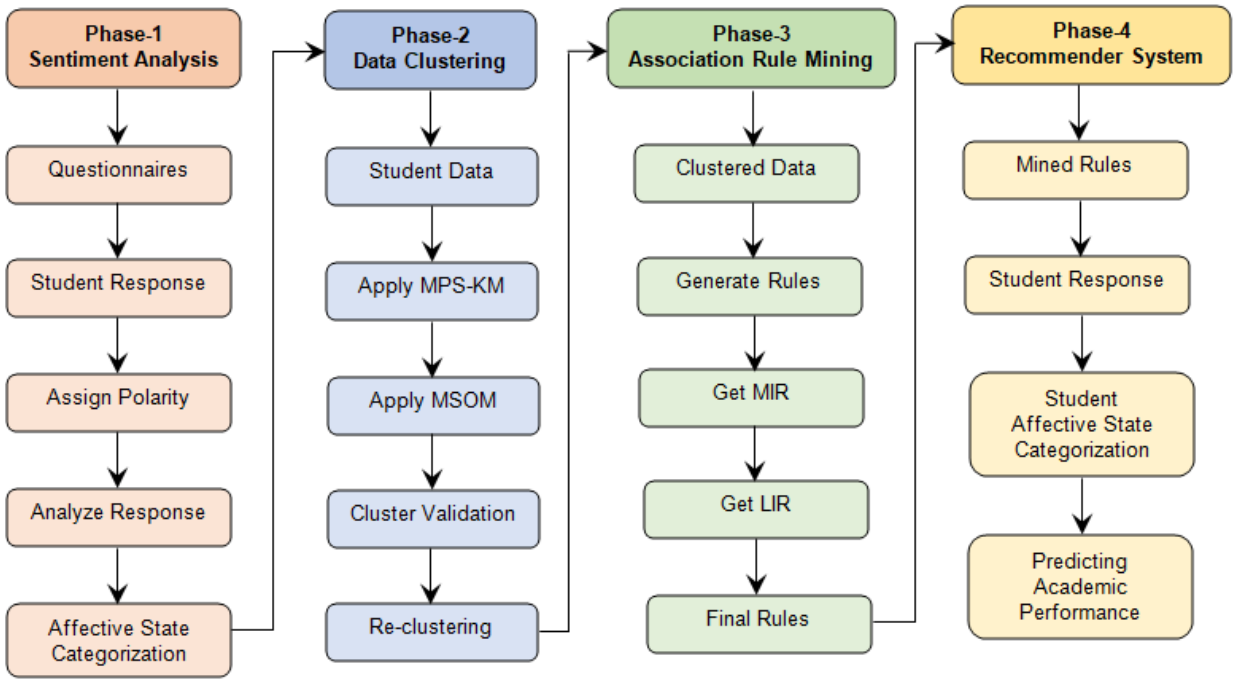

Figure 1 Workflow of Research work

Phase 1: Scoring Method for Closed-Ended Questionnaires Using Sentiment Analysis with Polarity value

Sentiment analysis identifies if a piece of content is positive, negative, or neutral. Sentiment analysis systems use natural language processing (NLP) and machine learning (ML) techniques to recognize, locate, and synthesize information and points of view from massive amounts of text. Sentiment analysis seeks to understand the public sentiment in a way that might

support company growth. This work uses sentiment analysis to analyze the students' affective states to help them academically. Open-ended and Close-ended questions are types of questionnaires. **Open-ended questions** are **questions** that allow for detailed, explanatory answers. Closed-ended questions are those that need respondents to select from a restricted set of pre-defined replies, such as "yes" or "no" or multiple-choice questions(Likert Scale, Rating Scale, Ranking, Checklist). Closed-ended questions are typically used to gather statistical information from the respondents. Closed-ended questionnaires to examine Students' Emotional Intelligence, Positive and Negative Emotions, Self Determination, self-esteem, Self-efficacy, and Emotional Happiness. This study suggests Quest SA, which analyses responses to a closed-ended questionnaire to determine the respective affective level of students using the sentiment analysis technique by assigning polarity scores to the options in the questionnaire. The sum of the polarity-value score is used to identify the students' affective categories.

Result: The polarity-based scoring method provides better MAE and accuracy than the standard scoring.

Inference: Based on closed-ended responses, this study creates a system for assessing questions and students' attitudes. A closed-ended questionnaire is used to collect the students' responses, and polarity-based sentiment analysis is used to determine the students' affective states.

Phase 2: An Efficient Clustering Algorithm for Mixed Data based on Re-clustering and Cluster Validation:

Clustering is an important strategy in data mining, which searches for groupings or clusters of data. In real-world applications, numerical and categorical features are routinely employed to define the data. Cluster validation is a technique used in clustering algorithms to assess the quality of the findings. Re-clustering and cluster validation are used in this study to offer an effective clustering approach for mixed numerical and category data. Two clustering techniques are initially used to cluster the mixed dataset: Modified Probability and Similarity-based K-Means (MPS-KM) clustering and Modified-Self-organizing Map (MSOM). Following the validation of the two algorithms, the optimal algorithm is chosen for re-clustering. It is an iterative process for enhancing the accuracy of cluster results. The wrongly clustered data is iteratively re-clustered and assessed based on the cluster validation.

Result: The proposed clustering model provides better performance metrics like Purity, Rand Index, Normalized Mutual Information, Precision, and Recall compared to other clustering approaches.

Inference: An efficient clustering method is suggested for grouping mixed numerical and categorical datasets. Cluster validation and iterative re-clustering improve the clustering performance. For many datasets, the recommended clustering approach was tested. The experiment findings validate the higher performance of the clustering algorithm.

Phase 3: ARMpred- An Efficient Prediction Based on Rule Generation for Clustered Data through Associative Classifier:

A rule-based model known as associative classification (AC) is used in data mining to create perfect classifiers using supervised learning data sets. It combines classification and association rule mining for effective classification. This work proposes an efficient prediction algorithm based on generated rules from clustered data through AC. The proposed approach contains three phases: Preprocessing, Rule Generation, and Prediction. In the first phase, the collected data is preprocessed using discretization. Then, it converts the numerical attribute to a categorical attribute. The number rules are generated in the second phase based on the cluster label. This work selects the More Important Rule (MIR) and Less Import Rule (LIR) for prediction. The final phase is a prediction that predicts the corresponding class based on generated rules. Based on the predicted class, it analyses the attributes in the ruleset.

Result: The proposed rule generation algorithm was evaluated using the number of rules, execution time, and accuracy. The proposed approach achieves high accuracy compared to other methods.

Inference: This work selects the More Important Rule (MIR) and Less Import Rule (LIR) for prediction based on two minimum support values. The proposed algorithm improves the accuracy. It also efficiently classifies the minority class label.

Phase4: Recommender System for Predicting Student academic performance with affective state Based on Closed-Ended Questionnaire Using Sentiment Analysis and Association Rule Mining:

Recommendation systems based on machine learning have evolved as a response to the overabundance of online information, which has proved problematic for users regarding the time spent on their searches and the quantity of information they retrieved. In educational recommendation systems, the usefulness of suggested educational materials will enhance the student's learning process. Thus it is crucial to adequately and reliably ensure valuable, relevant information. Depending on the recommendation method, most recommendation engines can be divided into three main subcategories: Collaborative filtering recommendation systems, content-based filtering recommendation systems, and hybrid recommendation systems. This work proposes a hybrid recommender system for predicting student academic performance by analyzing the affective state. The proposed recommender system is trained with the data processed by Quest-SA, Re-clust and rules mined by ARMpred. This work uses association rule mined to recommend student academic performance based on behavioural/emotional attribute

Result: Experiments with seven real-time questionnaires (Emotional intelligence, Eysenck personality, Self-determination scale, General Self-efficacy, Rosenberg's self-esteem, Positive and Negative affect schedule, and Oxford Happiness) are used to analyze the performance of the work.

Inference: This research work improves the student's personality. This is the first study of a recommender system for students based on closed-ended questionnaires.

Research Contributions

This research aims to develop a system for the student to identify the affective states and predict academic performance using sentiment analysis with proposed hybrid machine learning algorithms. The contribution of this research work is as follows:

1. Sentiment analysis is used to discover a student's affective state. The student responses are collected for the respective questionnaire, which is a close-ended questionnaire. The polarity value is assigned for each question in the questionnaire. The affective state scale is identified using sentiment analysis with polarity scores.

2. An efficient clustering algorithm is proposed for mixed numerical and categorical data based on re-clustering and cluster validation called Reclust. It is an iterative process for improving the quality of cluster results. The incorrectly clustered data is iteratively re-clustered and evaluated based on the cluster validation.
3. Using associative classification, a prediction algorithm is proposed to generate association rules from clustered data.
4. A recommendation system is suggested to predict the student's academic performance with the association rule generated.

Conclusion

In this research, a polarity score is assigned to the choices of the questionnaire. Based on the total score, students are categorized using sentiment analysis. The proposed Reclust clustering technique and ARMpred are employed to find an association between affective state and academic performance. The clustering is an effective technique to classify the student objects based on Emotional Intelligence Questionnaire, Rosenberg Self Esteem Scale, Self Determination scale, Positive Affect and Negative Affective scale, General Efficacy Scale, Oxford Happiness Inventory, Eysenck Personality questionnaire, and models each cluster respect to the questionnaire. The performance-based clusters are observed as Excellent, Very Good, Good, Above Average, Average, and Fail. This Experiment reveals the high academic competence of a high affective state and suggests the recommendation to concentrate on emotionally low students for Psychological change. The Correlation analysis divulges a strong association between Emotional Intelligence, Personality, Self Esteem, Emotions, Self Determination, Self Efficacy, Emotional Well-being and Academic Performance. The association rules were mined to design the recommender system to predict student candidates' academic performance during the learning process.

Scope for Further Research

In the future, our research aims to investigate and apply novel methodologies to predict students' academic performance concerning different affective parameters. This research uses a closed-ended questionnaire to predict the student's affective state. In the future, different questionnaire types will be used for student affective state prediction. The system can be integrated into an e-learning environment to predict the academic performance and analyze the effectiveness of e-learning. To predict student affective state and academic performance based on the impact of the institution/discipline and teaching.

LIST OF PUBLICATIONS

CONFERENCE ATTENDED AND PAPER PRESENTED:

1. Impact of Personality Traits on Students 'Academic Performance , presented in the International Conference on High Performance and Intelligent Computing organized by PSG college of Technology on Dec 9 2022.
2. A Study on the Impact of Self-Determination on Academic Achievement , presented in International Conference on Artificial Intelligence & Data Analytics, Internet of Things, Cyber Security (ICAIC'22)organized by Kumaraguru College of Technology, Coimbatore on May 21 & 22, 2022.
3. 3.ARM-pred- an efficient prediction based on rule generation for clustered data through associative classifier in the National Conference on Multidisciplinary Research for Sustainable Innovations (NCMRSI 2023) conducted by the Department of Computer Science, on 10th February 2023
4. Recommender System For Predicting Student Personality With Emotion Based On Closed-Ended Questionnaire Using Sentiment Analysis And Association Rule Mining in IEEE Conference on Application of AI insustainable computing- ITT2023 , Higher Colleges of Technology, United Arab Emirates.(Submitted)

JOURNALS PUBLISHED:

1. **“**Role Of Educational Data Mining in Student Learning Processes with Sentiment Analysis: A Survey”, International Journal Of Knowledge And Systems Science, Volume 11, Issue 4 October-December 2020.
2. “Quest_Sa: Pre-processing Method For Closed-Ended Questionnaire Using Sentiment Analysis Through Polarity”, Hindawi Mobile Information Systems Volume 2022, Article ID 4733550, 12 pages, https://doi.org/10.1155/2022/4733550 , September 2022.
3. "Reclust: An Efficient Clustering Algorithm For Mixed Data Based On Reclustering And Cluster Validation”, Indonesian Journal of Electrical Engineering And Computer Science.

REFERENCES

1. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining : Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of Business Research, 94, 335–343
2. Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artifcial intelligence methods to assess academic achievement in public high schools of a European Union country. Heliyon.
3. Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. Higher Education, 80(5), 875–894.
4. Hung, H. C., Liu, I. F., Liang, C. T., & Su, Y. S. (2020). Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. Symmetry, 12(2), 213.
5. De Medio, C., Limongelli, C., Sciarrone, F., & Temperini, M. (2020). MoodleREC: A recommendation system for creating courses using the moodle e-learning platform. Computers in Human Behavior, 104, 106168.
6. Baidada, M., Mansouri, K., & Poirier, F. (2022). Hybrid Filtering Recommendation System in an Educational Context: Experiment in Higher Education in Morocco. International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 17(1), 1-17.

7. Toçoğlu, M. A., & Onan, A. (2021). Sentiment analysis on students' evaluation of higher educational institutions. In Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference, Istanbul, Turkey, July 21-23, 2020 (pp. 1693-1700). Springer International Publishing.
8. Chen, W., Liu, B., Zhang, X., & Al-Qudah, I. (2022, January). An Event-Based Framework for Facilitating Real-Time Sentiment Analysis in Educational Contexts. In 2022 11th International Conference on Educational and Information Technology (ICEIT) (pp. 57-61). IEEE.
9. Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers' evaluation. *Applied computational intelligence and soft computing*, *2016*.
10. Nasim, Z., Rajput, Q., & Haider, S. (2017, July). Sentiment analysis of student feedback using machine learning and lexicon based approaches. In *2017 international conference on research and innovation in information systems (ICRIIS)* (pp. 1-6). IEEE.
11. Li, X., Wu, Z., Zhao, Z., Ding, F., & He, D. (2021). A mixed data clustering algorithm with noise-filtered distribution centroid and iterative weight adjustment strategy. *Information Sciences*, *577*, 697-721.
12. D'urso, P., & Massari, R. (2019). Fuzzy clustering of mixed data. *Information Sciences*, *505*, 513-534.
13. Rodriguez-Sanchez, F., Bielza, C., & Larrañaga, P. (2022). Multi-partition clustering of mixed data with Bayesian networks. *International Journal of Intelligent Systems*, *37*(3), 2188-2218.
14. Rajab, K. D. (2019). New associative classification method based on rule pruning for classification of datasets. *IEEE Access*, *7*, 157783-157795.
15. Pal, P. R., Pathak, P., & Luma-Osmani, S. (2021). IHAC: Incorporating Heuristics for Efficient Rule Generation & Rule Selection in Associative Classification. *Journal of Information & Knowledge Management*, *20*(01), 2150010.
16. Mattiev, J., & Kavšek, B. (2020). CMAC: clustering class association rules to form a compact and meaningful associative classifier. In *Machine Learning, Optimization, and Data Science: 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020*